Some Applications of Statistical Methods to the Analysis of Physical and Engineering Data

By W. A. SHEWHART

SYNOPSIS: Whenever we measure any physical quantity we customarily obtain as many different values as there are observations. From a consideration of these measurements we must determine the *most probable value;* we must find out *how much* an observation may be expected to vary from this most probable value; and we must learn as much as possible of the *reasons why* it varies in the particular way that it does. In other words, the real value of physical measurements lies in the fact that from them it is possible to determine something of the nature of the results to be expected if the series of observations is repeated. The best use can be made of the data if we can find from them the most probable frequency or occurrence of any observed magnitude of the physical quantity or, in other words, the most probable law of distribution.

It is customary practice in connection with physical and engineering measurements to assume that the arithmetic mean of the observations is the most probable value and that the frequency of occurrence of deviations from this mean is in accord with the Gaussian or normal law of error which lies at the foundation of the theory of errors. In most of those cases where the observed distributions of deviations have been compared with the theoretical ones based on the assumption of this law, it has been found highly improbable that the groups of observations could have arisen from systems of causes consistent with the normal law. Furthermore, even upon an a priori basis the normal law is a very limited case of a more generalized one.

Therefore, in order to find the probability of the occurrence of a deviation of a given magnitude, it is necessary in most instances to find the theoretical distribution which is more probable than that given by the normal law. The present paper deals with the application of elementary statistical methods for finding this *best* frequency distribution of the deviations. In other words, the present paper points out some of the limitations of the theory of errors, based upon the normal law, in the analysis of physical and engineering data; it suggests methods for overcoming these difficulties by basing the analysis upon a more generalized law of error; it reviews the methods for finding the best theoretical distribution and closes with a discussion of the magnitude of the advantages to be gained by either the physicist or the engineer from an application of the methods reviewed herein.

INTRODUCTION

WE ordinarily think of the physical and engineering sciences as being exact. In a majority of physical measurements this is practically true. It is possible to control the causes of variation so that the resultant deviations of the observations from their arithmetic mean are small in comparison therewith. In the theory of measurements we often refer to the "*true value*" of a physical quantity: observed deviations are considered to be produced by errors existing in the method of making the measurements.

With the introduction of the molecular theory and the theory of quanta, it has been necessary to modify some of our older conceptions. Thus, more and more we are led to consider the problem of measuring any physical quantity as that of establishing its most probable value. We are led to conceive of the physico-chemical laws as a statistical determinism to which "the law of great numbers" [1] imparts the appearance of infinite precision. In order to obtain a more comprehensive understanding of the laws of nature it is becoming more necessary to consider not only the average value but also the variations of the separate observations therefrom. As a result, the application of the theory of probabilities is receiving renewed impetus in the fields of physics and physical chemistry.

Statistical Nature of Certain Physical Problems. As typical of the newer type of physical problem, we may refer to certain data given by Prof. Rutherford and H. Geiger.[2] In this experiment the number of alpha particles striking, within a given interval, a screen subtending a fixed solid angle was counted. Two thousand six hundred and eight observations of this number were made. The first column of Table I records the number of alpha particles striking this screen within a given interval. The second column gives the frequency of occurrence corresponding to the different numbers in the first column.

TABLE 1

No. of Alpha Particles	Observed Frequency of Occurrence
0	57
1	203
2	383
3	525
4	532
5	408
6	273
7	139
8	45
9	27
10	10
11	4
12	0
13	1
14	1

It is obviously impossible from the nature of the experiment to attribute the variations in the observed numbers to errors of observation. Instead, the variations are inherent in the statistical nature of the phenomenon under observation.

[1] Each class of event eventually occurs in an apparently definite proportion of cases. The constancy of this proportion increases as the number of cases increases.

[2] *Philosophical Magazine*, October, 1910.

The questions which must be answered from a consideration of these data are typical. For example, we are interested to know how a second series of observations may be expected to differ if the same experiment were repeated. The largest observed frequency corresponds to four alpha particles, although what assurance is there that this is the most probable number? What is the probability that any given number of alpha particles will strike the screen in the same interval of time? Or again, what is the maximum number of alpha particles that may be expected to strike the screen? All of these questions naturally can be answered providing we can determine the most probable frequency distribution.

Statistical Nature of Certain Telephone Problems. The characteristics of some telephone equipment cannot be controlled within narrow limits much better than the distribution of alpha particles could be controlled in the above experiment. We shall confine our attention primarily to a single piece of equipment. The carbon microphone. For many reasons it is necessary to attain a picture of the way in which a microphone operates. It is necessary to find out why carbon is the best known microphonic material. In order to do this we must measure certain physical and chemical characteristics of the carbon and compare these with its microphonic properties when used under commercial conditions. In the second place it becomes necessary to establish methods for inspecting manufactured product in order to take account of any inherent variability, and yet not to overlook any evidence of a "trend" in the process of manufacture toward the production of a poor quality of apparatus. In the third place it so happens that the commercial measure of the degree of control exhibited in the manufacture of the apparatus must be interpreted ultimately in terms of sensation measures given by the human ear. That is, the first phase of the problem is purely physical; the second is one of manufacturing control and inspection and the third involves the study of a variable quantity by means of a method of measurement which in itself introduces large variations in the observations.

In one of the most widely used types of microphones there are approximately 50,000 granules of carbon per instrument. Each of these granules is irregular in contour, porous and of approximately the size of the head of a pin. If such a group of granules is placed in a cylindrical lavite chamber about $\frac{1}{2}$-inch in diameter and closed at either end with gold-plated electrodes; if this chamber is then placed on a suspension free from all building vibrations and carefully insulated from sound disturbances; if automatically controlled

mechanical means are provided for rolling this chamber at any desired speed; if all of the air and sorbed gases are removed from the carbon chamber and pure nitrogen is substituted; if the mean temperature is kept constant within 2° C; and if means are provided for measuring the resistance of the granules when at rest by observing the voltage across the two electrodes while current is allowed to flow for a period less than 1/200 of a second, it is found that the resistance (for most samples of carbon) may be determined within a fraction of one per

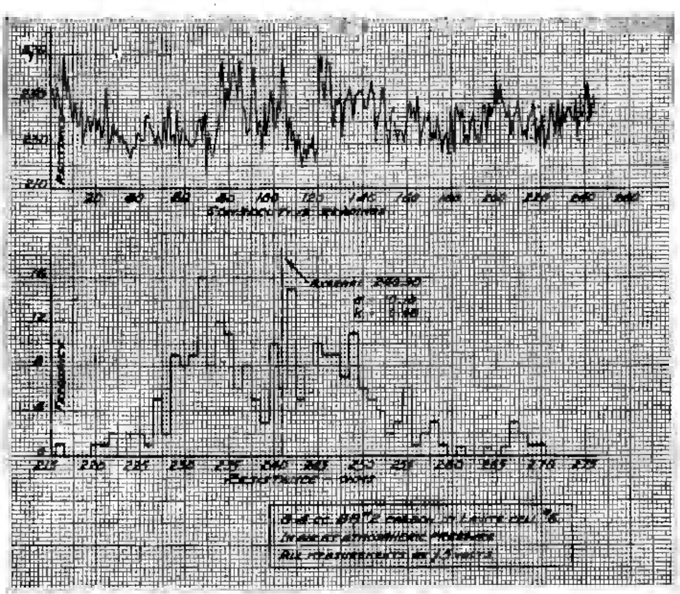


Fig. 1

cent. If, however, the button is rotated (even as slowly as possible) and then brought to rest, the resistance may differ several ohms from its first value. If a large number of observations are made after this fashion, we may expect to find for certain samples of carbon a set of values such as given in Fig. 1. The 270 observations of resistance reproduced in this figure were made on a sample of carbon at 1½ volts under conditions quite similar to those outlined above. The observed variation is from approximately 215 to 270 ohms. The upper curve is that of the resistance vs. the serial number of the readings. There is no apparent trend in the change of resistance from one reading to another. The lower curve in this figure shows the frequency histogram of the results. Attention is directed to the

wide variation in the observations, and to the fact that the frequency histogram appears to be bimodal.[3] Methods of dealing with such distributions will be considered.

Samples of carbon having different molecular surface structures have different resistances. To put it in a still more practical way, if the manufacturing process is not controlled within very narrow limits, wide variations are produced in the molecular properties of the carbon. The microphonic properties of these carbons are therefore different. One of the problems with which we have been concerned is to determine the relationship existing between the physical and chemical characteristics of the carbon and the resistance of the material when measured under different conditions. We are obviously dealing in this case with problems involving the measurement of physical quantities which cannot be controlled even in the labora-

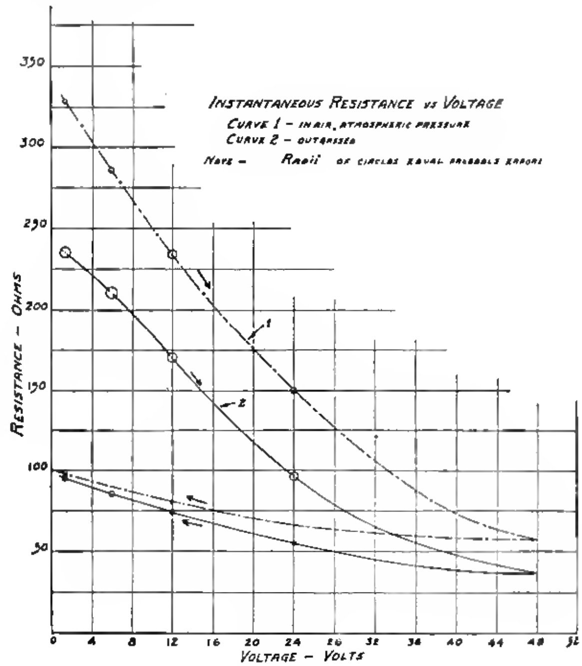


Fig. 2

[3] If curves which touch the axis at $+\infty$ and $-\infty$ have more than one value of the variable for which the derivative of the frequency in respect to the variable is equal to zero—the points being other than that for which the frequency is zero—these curves are referred to as bimodal, trimodal, etc. The modal value is the most probable one and is of particular interest in unimodal curves.

tory. If we remove the air and measure the resistance at different voltages, we may expect to find changes in the resistance similar to those indicated in Fig. 2. Curves 1 and 2 were taken for increasing voltages. The return curves were taken with decreasing voltage. Removal of the air from this particular sample of carbon produces comparatively large changes in the resistance. The resistance at 1½ volts is several times that at 48 volts. These curves were taken under conditions wherein all of the other factors were controlled. A sufficient number of observations was made in each case in order to establish the probable errors of the points as indicated by the radii of

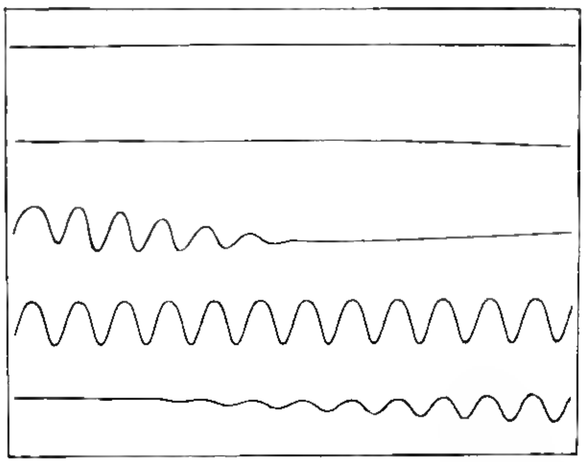


Fig. 3—Possible Types of Breathing of Granular Carbon Microphone.

the circles. If this same experiment were carried on at a different temperature, radically different results would be obtained.

If, instead of allowing the current to flow for a short interval of time, a continuous record is made of the resistance of the carbon while practically constant current flows through the carbon, the resistance will be found to vary. The maximum resistance reached in certain instances may amount to several times the minimum value. In general, this phenomenon is attributed to the effects of gas sorbed on the surface of the material. Transmitters cannot be made of lavite so that the expansions and contractions of the piece parts thereof augment the changes in resistance. This phenomenon, termed "breathing," may be, but seldom is, regular or periodic. An exceptional case of breathing is shown in Fig. 3. This was obtained with a special type of carbon in a commercial structure. The curves

themselves represent the current through the transmitter and, therefore, are inversely proportional to the resistance. All five curves were obtained with the same carbon in the same chamber by varying merely the configuration of the granules by slightly tapping the carbon chamber.

All of these effects can be modified to a large extent by varying the process of manufacture of the granular material. In practice it is necessary to know why slight changes in the manufacturing process cause large variations in the resistance characteristics of the carbon. The same process that improves one microphonic property may prove a detriment to another. It is in the solution of some of these problems that statistical methods have been found to be of great value in the interpretation of the results.

Whereas the physicist ordinarily works in the laboratory under controlled conditions, the engineer must work under commercial conditions where it is often impractical to secure the same degree of control. More than 1,500,000 transmitters are manufactured every year by the Bell System. Causes of variation other than those introduced by the carbon help to control the transmitter. For example, variations may be introduced by the process of assembly, or by differences in the piece parts of the assembled instrument. The measure of the faithfulness and efficiency of reproduction depends fundamentally upon the human ear. Obviously all transmitters cannot be tested. Instead, we must choose a number of instruments and from observations made on these determine whether or not there is any trend in the manufactured product. Naturally we may expect to find certain variations in the results according to the rules of chance. To take the simplest illustration, we may flip a coin 6 times. Even if it is symmetrical we may expect occasionally to find all heads and occasionally all tails, although the most probable combination is that of 3 heads and 3 tails. We must, therefore, determine first of all whether or not the observed variations are consistent with those due to sampling according to the laws of chance. If there is an apparent trend in product, the data should be analyzed in order to determine, if possible, whether it is due to lack of control in the manufacture of carbon or to some other set of causes such as mentioned above. Because of economic reasons we must keep the number of observations at a minimum consistent with a satisfactory control of the product. Here again it has been found that the application of statistical methods is necessary to the solution of the problems involved.

Before considering the problem of the measurement of efficiency and quality of the transmitter, let us consider the schematic diagram

of the telephone system as shown in Fig. 4. Essentially this consists of the transmitter, the line and the receiver. The oldest method of measurement is to compare one transmitter against a standard in the following way. An observer calls first in the standard and then in the test transmitter, while another observer at the receiving end judges the faithfulness of reproduction. The pressure wave striking the transmitter diaphragm varies with the observer and also with the degree of mechanical coupling between the sound source

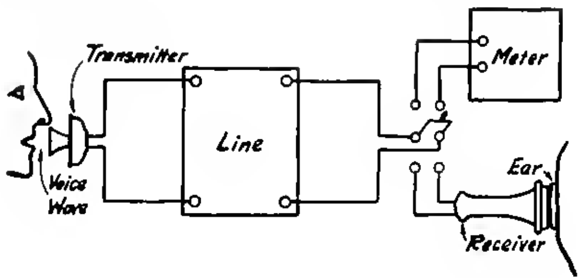


Fig. 4

and the diaphragm of the instrument. The judgment of the observer at the receiving end in influenced by physiological and psychological causes. Obviously it is desirable that such a method be supplanted by a machine test which will eliminate the variabilities in the sound source and in the human ear. Up to the present time the nature of speech and the characteristics of the human ear are not known sufficiently well to establish either an ideal sound source or an electrical meter to replace the human voice and ear respectively. The best that can be done is to approximate this condition. Even though the meter readings may be the same, the simultaneous observations made with the ear in general will be different. A calibration of the machine must, therefore, depend upon a study of the degree of correlation between the average measure given by the machine and that given by the older method of test.

Thus, we see how special problems arise in the fields of both physics and engineering wherein it is impossible to control the variations. In what way, if any, are these problems related, or is it necessary to attack each one in a different manner? We shall see that all of these problems are in a way fundamentally the same and that the same method of solution can be applied to all of them. This is true because it is necessary to determine in every instance the law of distribution of the variable about some mean value.

Why Do We Need to Know the Law of Deviation of the Different Observations About Some Mean Value?

In all of the above problems as in every physical and engineering one, certain typical questions arise which can be answered only if we know the law of distribution $y=f(x)$ of the observations where y represents the frequency of occurrence of the deviations x from some mean value. At least three of these questions are the same for both fields of investigation.[4]

Let us consider the physical problem. From a group of n observations of the magnitude of a physical quantity, we obtain in general n distinct values which can be represented by $X_1, X_2, \ldots X_n$. From a study of these we must answer the following questions:

1. What is the most probable value?
2. What is the frequency of occurrence of values within any two limits?
3. Is the set of observations consistent with the assumption of a random system of causes?

The answers to these questions are necessary for the interpretation of Prof. Rutherford's data referred to above: They are required in order to interpret the data presented in Fig. 1 which are typical of physical and chemical problems arising in carbon study; these same answers are fundamentally required in the analysis of all physical data. These questions can be answered from a study of the frequency distribution. If this be true, it is obvious that the statistical methods of finding the best distribution are of interest to the physicist.

Let us next consider the engineering problem where we shall see that the same questions recur. Assuming that manufacturing methods are established to produce a definite number of instruments within a fixed period, one or more of the characteristics of these instruments must be controlled. We may represent any one of these characteristics by the symbol X. The total number of instruments that will be manufactured is usually very indefinite. It is, however, always finite. Even with extreme care some variations in the methods of manufacture may be expected which will produce

[4] In order to calibrate the machine referred to in a preceding paragraph and also to determine the relationships between the physico-chemical and microphonic properties of carbon, it was necessary to study the correlation between two or more variables, but in each case it was necessary to determine first the law of distribution for each variable in order to interpret the physical significance of the measures of correlation because this depends upon the laws of distribution. The reason for this is not discussed in the present paper, for attention is here confined to the method of establishing the best theoretical frequency distribution derived from a study of the observations.

variations from instrument to instrument in the quantity X. After the manufacturing methods have been established, the first problem is to obtain answers to the following questions:

1. What is the most probable value of X?
2. What is the percentage of instruments having values of X between any two limits?
3. Are the causes controlling the product random, or are thev correlated? [5]

In this practical case we must decide to choose a certain number of instruments in order to obtain the answers to these questions; that is, to obtain the most probable frequency distribution. We must, however, go one step further. We must choose a certain number of instruments at stated periods in order to determine whether or not the product is changing. How big a sample shall we choose in the first place, and how large shall the periodic samples be? Obviously it is of great economic importance to keep the sample number in any case at a minimum required to establish within the required degree of precision the answers to the questions raised.

The close similarity between the physical and engineering problems must be obvious. Naturally, then, we need not confine ourselves in the present discussion to a consideration of only the problems arising in connection with the study of those microphonic properties of carbon which gave rise to the present investigation. Several examples are therefore chosen from fields other than carbon study. However only those points which have been found of practical advantage in connection with the analysis of more than 500,000 observations will be considered.

The type of inspection problem may be illustrated by the data given in Table II.

The symbol X refers to the efficiency of transmitters as determined in the process of inspection: N represents the number of instruments measured in order to obtain the average value X. The first four rows of data represent the results obtained by four inspection groups G_1, G_2, G_3 and G_4. The results given are for the same period of time. The next three rows are those for different machines M_1, M_2 and M_3. The last row gives the results of single tests on 68,502 transmitters, a part of which was measured on each of the three machines. The third column in the table gives the standard deviations. It will be observed

[5] The significance of this question will become more evident in the course of the paper. We shall find that, if the causes are such as to be technically termed random, we can answer all practical questions with a far greater degree of precision than we can if the causes are not random.

TABLE II

INSPECTION DATA ON TRANSMITTERS

	$\overline{X}$	σ	$\frac{3\sigma}{\sqrt{N}}$	N	k	σ_k	β_2	σ_{β_2}	$\sigma_{\overline{X}}$	$3\sigma_k$	$3\sigma_{\beta_2}$	Pearson Type
G_1	.548	.739	.0131	4510	− .214	.056	4.152	.073	.011	.108	.219	IV
G_2	.740	.896	.0533	2540	− .949	.049	4.426	.097	.018	.147	.291	VI
G_3	.766	.762	.0568	1620	− .109	.061	5.176	.122	.019	.183	.366	VI
G_4	.934	.677	.0398	2610	−1.413	.048	7.677	.096	.013	.144	.288	IV
M_1	−1.66	1.32	.0386	10855	− .70	.024	3.128	.047	.013	.072	.141	
M_2	−1.69	1.07	.0300	11577	− .84	.023	4.240	.046	.010	.069	.138	
M_3	−1.79	1.04	.0510	3749	− .56	.040	3.628	.080	.017	.120	.240	
Machines 1, 2, 3	−1.641	1.14	.0131	68502	− .80	.009	Out		.004	.027		I

that comparatively large differences exist between the averages obtained for different groups of transmitters by different groups of observers. Similarly, comparatively large variations exist in these averages even when taken by the machines (the large difference between the sensation and machine measures is due to a difference in the standard used, corrections for which are not made in this table).

Are these differences significant? Is product changing? That is, are the manufacturing methods being adequately controlled? Are these results consistent with a random variation in the causes controlling manufacture? These are the questions that were raised in connection with the interpretation of these data. The ordinary theory of errors gives us the following answer. It will be recalled that the standard deviation (or the root mean square deviation) of the average $\sigma_{\overline{X}}$ is equal to $\frac{\sigma}{\sqrt{N}}$. Also, from the table of the normal probability integral we find that the fractional parts of the area within certain ranges are as follows: For the ranges $\overline{X} \pm \sigma$, $\overline{X} \pm 2\sigma$, and $\overline{X} \pm 3\sigma$, we have the percentages 68.268, 95.450, and 99.730 respectively. Obviously, it is highly improbable that the difference between averages should be greater than three times the standard deviation of the average, providing we assume that all of the samples were drawn from the same universe: In other words, that all of the samples were manufactured under the same random conditions. The fourth column, then, indicates practical limits to the variations in the averages. It is obvious, therefore, that the differences between the averages are larger than could have been expected, if the same system of causes controlled the different groups of observations. In other words the differences are significant and must be explained.

Why do these variations exist? We shall show in the course of the discussion that the normal law is not sufficient to answer these questions. We shall show also that the variations noted are largely the result of the method of sampling used at that time. The significance of the other factors given in this table is discussed later.

Why Is the Application of the Normal Law Limited?

Why can we not assume that the deviations follow the normal law of error? This is

$$y=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{x^2}{2\sigma^2}} \tag{1}$$

where σ is the root mean square error $\sqrt{\frac{\Sigma yx^2}{n}}$ and y is the frequency of occurrence of the deviation x from the arithmetic mean and n is the number of observations? If they do, the answers to all of the questions raised in the preceding paragraphs can be easily answered in a way which is familiar to all acquainted with the ordinary theory of errors and the method of least squares. This is an old and much debated question in the realm of statistics. Let us review briefly some of the a posteriori and a priori reasons why the normal law has gained such favor and yet why it is one of the most limited, instead of the most general, of the possible laws.

A Posteriori Reasons. The original method of explaining the normal law rests upon the assumption that the arithmetic mean value of the observations is always the most probable. Since experience shows that the observed arithmetic mean seldom satisfies the condition of being the most probable we may justly question the law based upon an apparently unjustified assumption.

Gauss first enunciated this law which is often called by his name. The fact that so great a mathematician proposed it led many to accept it. He assumes that the frequency of occurrence of a given error is a function of the error. The probability that a given set of n observations will occur is the product of the probabilities of the n independent events. He then assumes that the arithmetic mean is the most probable and finds the equation of the normal law. Thus he *assumes* the answer to the first question; that is, he assumes that the most probable value is always the arithmetic mean. In most physical and engineering measurements the deviations from the arithmetic mean are small, and the number of observations is not sufficiently large to determine whether or not they are consistent

with the assumption of the normal law. Under these conditions this law is perhaps as good an approximation as any.

The fundamental assumptions underlying the original explanation were later brought into question. What a priori reason is there for assuming that the arithmetic mean is the most probable value? Why not choose some other mean? [6] Thus if we assume that the median [7] value is the most probable, we obtain as a special case the law of error represented by the following equation:

$$y = Ae^{-h^2|x|} \tag{2}$$

where y represents the frequency of occurrence of the deviation x from the median value and e is the Naperian base of logarithms. Both A and h are constants. If, however, we assume that the geometric mean is the most probable, we have as a special case the law of error represented by the following equation:

$$y = Ae^{-h^2(\log X - \log a)^2} \tag{3}$$

where in this case y is the frequency of occurrence of an observation of magnitude X, "a" is the true value, and A and h are constants.[8]

Enough has been said to indicate the significance of the assumption that the arithmetic mean is the most probable value, but, why choose this instead of some other mean? No satisfactory answer is available. So far as the author has been able to discover, no distribution representing physical data has even been found which approaches the median law. Several examples have been found in the study of carbon which conform to the law of error derived upon the assumption that the geometric mean is the most probable. If the arithmetic mean were observed to be the most probable in a majority of cases, we might consider this an a posteriori reason for accepting the normal law. We find the contrary to be the case.

Furthermore, we find in general that the distribution of errors is non-symmetrical about the mean value. In fact, most of the distributions which are given in textbooks dealing with the theory of errors and the method of least squares to illustrate the universality

[6] An average or mean value may be defined as a quantity derived from a given set of observations by a process such that if the observations became all equal, the average will coincide with the observations, and if the observations are not all equal, the average is greater than the least and less than the greatest.

[7] If a series of n observations are arranged in ascending order of magnitude, the median value is that corresponding to the observation occurring midway between the two ends of the series.

[8] A very interesting discussion of the various laws that may be obtained by assuming different mean values is given in J. M. Keynes' "A Treatise on the Theory of Probability."

of the law are, themselves, inconsistent with the assumption of such a law. Prof. Pearson was one of the first to point out this fact. He considers among others an example originally given by Merriman [9] in which the observed distribution is that of 1,000 shots fired at a target. The theoretical normal is the solid line in Fig. 5 and the

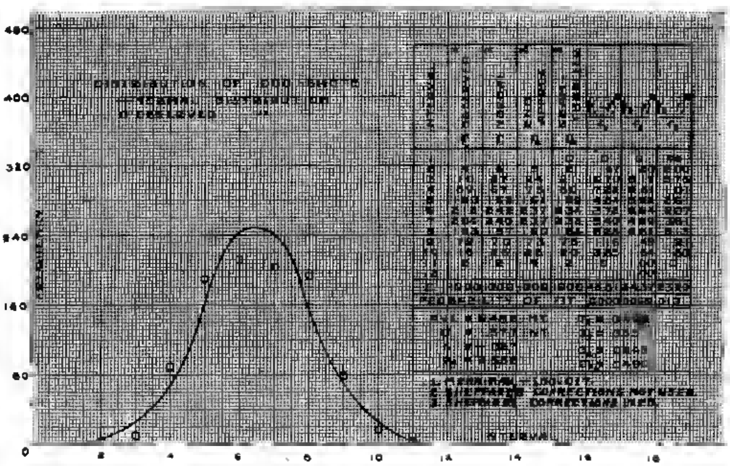


Fig. 5

observed frequencies are the small circles. When represented in this way there appears to be a wide divergence between theory and experience. Of course, some divergence may always be expected as a result of variations due to sampling; and, too, we must always question a judgment based entirely upon visual observation [10] of a graphical representation of this character. Prof. Pearson uses his method—which will be discussed later—for measuring the goodness of fit between the theoretical and observed distributions. He [11] finds that a fit as bad or worse than that observed could have been expected to occur on an average of only 15 to 16 times in ten million. We must conclude, therefore, that these data are not consistent with the assumption of a universal normal law.

A Priori Reasons. From the physicist's viewpoint the origin of the Gaussian law may be explained upon a more satisfactory basis.

[9] "Method of Least Squares," Eighth Edition—Page 14.

[10] This point will be emphasized later:—first, by showing that these data *appear* consistent with a normal law when plotted on probability paper, and second, by showing that some frequency distributions appear normal when plotted even though they are not. The other data in this table will be referred to later.

[11] Reference to the original article and a quotation therefrom given in the eleventh edition of the *Encyclopedia Britannica* on the article "Probability."

It is that which was originally suggested by La Place. If, however, we accept this explanation, we must accept the fact that the normal law is the exception and not the rule. Let us consider why this is true.[12]

This method of explanation rests upon the assumption that the normal law is the first approximation to the frequencies with which different values will be assumed by a variable quantity whose variations are controlled by a large number of independent causes acting in random fashion. Let us assume that:

a. The resultant variation is produced by n causes.
b. The probability p that a single cause will produce an effect Δx is the same for all of the causes.
c. The effect Δx is the same for all of the causes.
d. The causes operate independently one of the other.

Under these assumptions the frequency distribution of deviations of 0, 1, 2 . . . n positive increments can be represented by the successive terms of the point binomial $N(q+p)^n$ where N represents the total number of observations.

Under these conditions if $p=q$ and $n=\infty$, the ordinates of the binominal expansion can be closely approximated by a *normal* curve having the same standard deviation. These restrictions are indeed narrow. In practice it is probable that p is never equal to q, and it is certain that n is never infinite. Therefore, the normal distribution should be the exception and not the rule.

There is a more fundamental reason, however, why we should seldom expect to find an observed distribution which is consistent with the normal law. In what has preceded we have assumed that each cause produced the same effect Δx, and that the total effect in any instance is proportional to the number of successes.

Let us assume that the resultant effect is, in general, a function of the number n of causes producing positive effects, that is, let $X=\phi(n)$. Thus we assume that the frequency distributions of the number of causes and of the occurrence of a magnitude X are respectively

$$y=f(n)$$

and

$$y_1=f_1\ (X)$$

for two values of n, say n and $n+dn$, there will be two values of X, say X and $X+dX$. The number of observations within this interval of n must be the same as that within the corresponding interval of X.

[12] Bowley "Elements of Statistics," Part II.

If the distribution in X is normal such that we have

$$y_1 = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(X-a)^2}{2\sigma^2}},$$

then

$$y = \frac{1}{\sigma\sqrt{2\pi}} \phi'(n) e^{-\frac{[\phi(n)-a]^2}{2\sigma^2}} \tag{4}$$

where a is the arithmetic mean value, therefore, the distribution of the causes need not be normal; conversely if the causes are distributed normally, the observations will not in general be normal.

This idea is of great importance in the interpretation of observed distributions of physical data.[13] To illustrate, let us assume that the natural causes which affect the growth of apples on a given tree produce a normal variation in the diameters of the apples. Obviously, the distribution of either the cross-sectional areas or the volumes will not be normal.[14]. If the distribution of the diameters is normal as supposed, the arithmetic means of these diameters is the most probable value. Obviously, however, neither the arithmetic mean area nor the arithmetic mean volume will be the most probable, because in general

$$\frac{1}{n}\Sigma f(X) \neq f\left(\frac{1}{n}\Sigma X\right). \tag{5}$$

As already indicated, the deviations dealt with in the present investigation were not small. The form of the observed distribution may be expected, therefore, to depend upon the functional relationship between the observed quantity and the number of causes. We shall

[13] Kapteyn, J. C.—Skew Frequency Curves—Groningen, 1903.

[14] In the theory of errors this fact is taken into account by assuming that the variations are always *small*. Thus, if the variable X can be represented as a function F of certain other variables $U_1, U_2, \ldots U_m$ so that we have

$$X = F(U_1, U_2, \ldots U_m),$$

we ordinarily assume that we can write this expression in the following form

$$X = F(a_1 + u_1, a_2 + u_2, \ldots a_m + u_m).$$

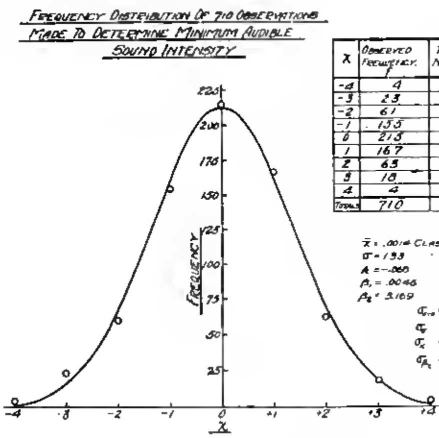
A further assumption is made that the u's are small so that 2nd and higher powers and products of these can be neglected. Under these conditions the distribution of X is normal and has a standard deviation given by the following expression:

$$\sigma_X = \sqrt{\left(\sigma_{U_1}\frac{\partial F}{\partial U_1}\right)^2 + \left(\sigma_{U_2}\frac{\partial F}{\partial U_2}\right)^2 + \cdots \left(\sigma_{U_m}\frac{\partial F}{\partial U_m}\right)^2}.$$

But, thus, we are led to overlook the significance of the form of F, particularly in those practical cases such as are of interest in the present paper where the quantities $u_1, u_2, \ldots u_m$ are not small.

illustrate the significance of these ideas as an aid in the interpretation of data by reference to the results of our study of the law of error of the human ear in measuring the efficiency of transmitters.

Let us consider the problem of determining the minimum audible sound intensity. Let us assume that there are n physiological and psychological causes controlling this sensation measure, and that the probabilities of the causes producing 0, 1, 2 . . . n effects are dis-



X	OBSERVED FREQUENCY f	THEORETICAL FREQUENCY NORMAL LAW f_1	THEORETICAL FREQUENCY 2ND APPROXIM. f_2	$\frac{(f-f_1)^2}{f_1}$	$\frac{(f-f_2)^2}{f_2}$
-4	4	3	3	.33	.33
-3	23	18	19	1.39	.84
-2	61	71	70	1.41	1.16
-1	155	159	156	.10	.01
0	215	208	208	.24	.24
1	167	159	162	.40	.15
2	65	71	72	.90	1.12
3	18	18	18	0	.00
4	4	3	2	.33	2.00
TOTALS	710	710	710	5.10	5.86
			P	.7463	.6636

$\bar{x}$ = .0014 CLASS INTERVALS = .0798 NAPIERS
σ = 1.33 " " = .7481
k = -.060
β_1 = .0046
β_2 = 3.169

$\sigma_{\bar{x}}$ = .0265
σ_σ = .0201
σ_k = .0919
σ_{β_2} = .180

Fig. 6

tributed normally. Because of these differences in human ears different amounts of sound energy are required to produce minimum audible sensations. What is the distribution of energies?

The data are given in Fig. 6. These have been previously reported by Fletcher and Wegel of this laboratory.[15] The method of making these measurements was described in their original papers. It is sufficient to recall that the results are given in terms of pressures in dynes per square centimeter. Seven hundred and ten observations covering the frequency range of from 60 to 8,000 cycles are included. The data include results for both ears of 14 women and 20 men, and one ear only for two women and two men. Only ears that had been medically inspected as being physiologically normal were selected. These results, therefore, include variations in the observations of a single observer with those of different observers.

The natural logarithms of the intensities were added and the average of these was obtained. The distribution of the natural

[15] Fletcher, H. and Wegel, R. L.—Proceedings of the National Academy of Science—Vol. VIII, pp. 5—6, January, 1922.
Physical Review, Vol. XIX, pp. 550 seq. 1922.

logarithms of the intensities is given in the second column of the table in Fig. 6. The smooth line is the normal curve based upon the observed value of standard deviation. The distribution of the logarithms of the intensities is normal.[16] The arithmetic mean of the logarithms is the most probable. Therefore, the distribution of intensities is decidedly skew, and the geometric mean intensity is the most probable. Here, then, is an excellent example in which it is highly probable that the distribution of the causes is random and normal, but in which the resultant effect is not a linear function of the number of causes.[17]

Can We Ever Expect to Find a Normal Distribution in Nature?

The answer is affirmative. If the resultant effect of the independent causes is proportional to their number, the distribution rapidly approaches normality as the number of causes is increased even though $p \neq q$.

To show this, let us assume that the variation in a physical quantity is produced by 100 causes, and that each cause produces the same effect Δx. Also, let us assume the probability p to be 0.1, that each cause produces a positive effect. The distribution of 0, 1, 2, . . . n successes in 1000 trials is given by the terms of the expansion 1000 $(.9+.1)^{100}$. Obviously such a distribution is skew, p is certainly not equal to q, and n is far from being infinite. If the normal law

[16] In fact this is an exceptionally close approximation to the normal law. This will be more evident after we have considered the methods for measuring the goodness of fit as indicated by the other calculations given in this figure. For the present it is sufficient to know that approximately 75 times out of 100 we must expect to get a system of observations which differ as much or more from the theoretical distribution calculated from the normal law than the observed distribution differs therefrom in this case. The fact that the second approximation does not fit the observed distribution as well as the normal—*i.e.* the measure of probability of fit P is less—indicates that the observed value of the skewness k is not significant.

[17] These results are of particular interest to telephone engineers. The fact that the distribution of the logarithms of the intensities is normal is consistent with the assumption of Fechner's law which states that the sensation is proportional to the logarithm of the stimulus. The range of variation (that is, $\bar{X} \pm 3\sigma$) in different observers' estimates of the sound intensity required to produce the minimum audible sensation is approximately 20 miles. The range of error of estimate depends upon the intensity of sound and decreases as the sound energy level increases. Thus for the average level which prevails for transmission over the present form of telephone system in a three mile loop common battery circuit it is less than 9 miles. Even at this intensity, however, it is obvious that although scarcely any observers will differ in their estimates by more than 9 miles, 50% of them will differ by at least 2 miles. These results also furnish experimental basis for the statement made in the beginning of this paper: that is, the variations introduced in the method of measurement of transmitter efficiencies are large in comparison with the average efficiency.

TABLE III—COMPARSION OF THE TERMS OF THE EXPANSION $1000(.1+.9)^{100}$ WITH THOSE OBTAINED BY VARIOUS APPROXIMATIONS

Number of Successes	* $1000(.1+.9)^{100}$ f	Normal Law f_1	2nd Approximation f_2	*Law of Small Numbers f_3	*Gram Charlier f_4	*Poisson Charlier f_5	† Pearson f_6	$\frac{(f-f_1)^2}{f_1}$	$\frac{(f-f_2)^2}{f_2}$	$\frac{(f-f_3)^2}{f_3}$	$\frac{(f-f_4)^2}{f_4}$	$\frac{(f-f_5)^2}{f_5}$	$\frac{(f-f_6)^2}{f_6}$
−1	.0	.2	.0	.0	.0	.0	.0	.200	.0	.0	.0	.0	.0
0	.1	.6	.0	.0	.0	.0	.9	.417	∞	∞	∞	∞	.711
1	.3	1.5	.4	.5	.4	.2	2.2	.960	.025	.080	.025	.050	1.641
2	1.6	3.9	2.1	2.3	2.0	1.6	7.1	1.356	.119	.213	.080	.0	4 261
3	5.9	8.9	6.7	7.6	6.5	5.9	15.9	1.011	.096	.380	.055	.0	6.289
4	15.9	18.3	16.7	18.9	16.2	15.9	28.9	.315	.038	.476	.006	.0	5.848
5	33.9	33.4	34.0	37.8	33.3	34.0	44.8	.007	.0	.402	.011	.0	2.652
6	59.6	54.8	58.8	63.0	58.1	59.9	61.3	.420	.011	.183	.039	.002	.047
7	88.9	80.9	88.0	90.1	87.5	89.3	76.1	.791	.009	.016	.022	.002	2.153
8	114.8	106.0	113.9	112.5	114.5	114.9	87.0	.731	.007	.047	.001	.0	8.883
9	130.4	125.2	130.5	125.1	131.8	130.1	92.8	.216	.0	.225	.015	.001	15.234
10	131.9	132.6	132.6	125.1	133.9	131.4	93.3	.004	.004	.370	.030	.002	15.970
11	119.9	125.2	119.9	113.7	121.1	119.4	89.1	.224	.0	.338	.012	.002	10.647
12	98.8	106.0	98.1	94.8	98.4	98.6	81.2	.489	.005	.169	.002	.0	3.815
13	74.3	80.9	73.8	72.9	73.2	74.4	71.0	.538	.003	.027	.017	.0	.153
14	51.3	54.8	50.8	52.1	50.2	51.4	59.9	.224	.005	.012	.024	.0	1.235
15	32.7	33.4	32.8	34.7	32.2	33.0	48.7	.015	.0	.115	.008	.003	5.257
16	19.3	18.3	19.9	21.7	19.4	19.7	34.5	.055	.018	.265	.001	.008	6.697
17	10.6	8.9	11.1	12.8	10.9	10.8	29.5	.325	.023	.378	.008	.004	12.109
18	5.4	3.9	5.7	7.1	5.8	5.5	22.0	.577	.016	.407	.028	.002	12.525
19	2.6	1.5	2.6	3.7	2.7	2.5	16.0	.807	.0	.327	.004	.004	11.223
20	1.2	.6	1.3	1.9	1.2	1.1	11.4	.600	.008	.258	.0	.009	9.126
21	.5	.1	.4	.9	.5	.5	7.9	1.600	.025	.178	.0	.0	6.932
22	.2	.1	.2	.4	.2	.1	5.4	.100	.0	.100	.0	.1	5.007
23	.0	.0	.0	.2	.0	.0	3.6			.200			3.600
24	.0	.0	.0	.0	.0	.0	2.3						2.300
25	.0	.0	.0	.0	.0	.0	1.5						1.500
26	.0	.0	.0	.0	.0	.0	.9						.900
27	.0	.0	.0	.0	.0	.0	.6						.600
28	.0	.0	.0	.0	.0	.0	.4						.400
29	.0	.0	.0	.0	.0	.0	.2						.200
30	.0	.0	.0	.0	.0	.0	.1						.100
Σ	1000.1	1000.0	1000.3	999.8	1000.0	1000.2	996.5	11.982	.412	5.166	.388	.189	158.015
Ave.	9.998	10.000	9.996	9.999	9.998	10.000	10.754						

$p=.1$ $\qquad \sigma=\sqrt{pqn}=3$ $\qquad \beta_2=3+\frac{1-6\,pq}{pqn}=3.511$ $\qquad \sigma_{\beta_2}=\sqrt{\frac{24}{1000}}=.155$

$q=.9$
$n=100$ $\qquad K=\frac{q-p}{\sqrt{pqn}}=.267$ $\qquad \sigma_k=\sqrt{\frac{6}{1000}}=.077$
$pn=10$

* Fisher, A.—"The Mathematical Theory of Probabilities." † Pearson, Karl—Phil. Mag., 1907, pp. 365–378.

were fitted to such a distribution, would it be possible to detect easily any great difference between theory and observation?

Let us compare the two distributions. The data are given in Table III. First, the average value must be the most probable in order to be consistent with the normal law. It is, because the observed most probable value corresponds to 10 successes, and the average of

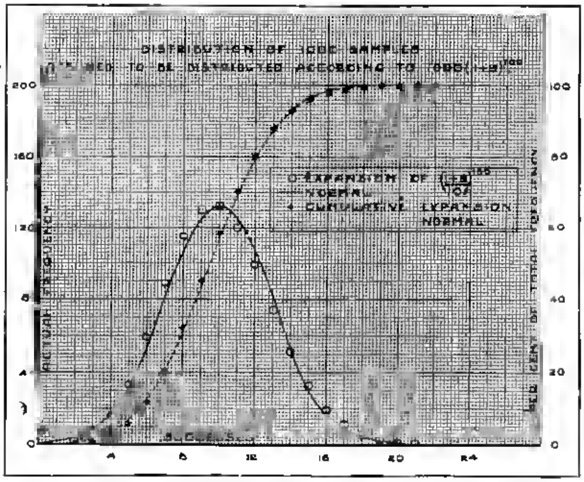


Fig. 7

the hypothetically observed distribution is 9.998. This under ordinary circumstances would be considered a close check between theory and practice.

The normal distribution is given in the third column of the table. Even though there is a difference between the frequencies given in the second and third columns, would the average observer be apt to conclude that the hypothetically observed distribution is other than normal? He would probably base his answer upon a graphical comparison such as given in Fig. 7. The solid line represents the normal curve; whereas the frequencies given in the second column of Table III are represented by circles. It is obvious that the normal law appears to be a very close approximation to the terms of the binomial expansion.

Thus we see that for even a small number of causes the difference between p and q may be quite large, and yet the difference between the distributions given by the binomial expansion and that given by the normal law is apparently small and not easily to be detected by ordinary methods. As n increases the closeness of fit does likewise.

If p is equal to q, the number of causes must be very small indeed before we are able to detect the difference between the terms of the binomial expansion and those given by the normal law. To show that this is true I have chosen a case corresponding to a physical condition where there are only 16 causes and where p is equal to q. The data are given in Table IV.

TABLE IV

Successes	$(.5+.5)^{16}$ f	Normal Law with same σ f_1
0	.0000153	.0000669
1	.0002441	.0004363
2	.0018311	.0022159
3	.0085449	.0087641
4	.0277710	.0269955
5	.0666504	.0647586
6	.1220825	.1209853
7	.1745605	.1760326
8	.1963806	.1994711
9	.1745605	.1760326
10	.1220825	.1209853
11	.0666504	.0647588
12	.0277710	.0269955
13	.0085449	.0087641
14	.0018311	.0022159
15	.0002441	.0004363
16	.0000153	.0000669

Obviously, therefore, the limitations imposed by the assumptions as to the number of causes and the equality of p and q are not as important as they might at first appear. It is probable that this is one of the reasons why we find approximately normal distributions. If, however, p is sufficiently small, the difference between the observed distribution and that consistent with the normal law can easily be detected. We shall show in a later section that this is true for Rutherford's data.[18]

Is There a Universal Law of Error ?

Obviously from what has already been said, the normal law is not a universal law of nature. It is probable that no such law exists. We do, however, have certain laws which are more general than the normal. We shall consider briefly some of these types in an effort to indicate the advantages that can be gained by an application of them to physical data.

[18] Loc. cit.

Binomial Expansion $(p+q)^n$. We have already seen that the distribution is approximately normal when $p=q$ and $n \doteq \infty$. Following Edgeworth[19], Bowley [20] shows that if $p \neq q$ but $n \doteq \infty$ the frequency y of the occurrence of a deviation of magnitude x is given by the following expression where k represents the skewness [21] of the distribution:

$$y=\frac{1}{\sigma\sqrt{2\pi}}\left(\exp. -\frac{x^2}{2\sigma^2}\right)\left[1-\frac{k}{2}\left(\frac{x}{\sigma}-\frac{x^3}{3\sigma^3}\right)\right]. \tag{6}$$

This will be referred to as the *second approximation.*

If p is very small, but $pn=\lambda$ is finite, we have the so-called *law of small numbers* [22] which was first derived by Poisson. The successive terms of the series $e^{-\lambda}\left(1+\lambda+\frac{\lambda^2}{2}+\frac{\lambda^3}{3}+\ldots\right)$ represent the chances of 0, 1, 2 . . . n successes. Theoretically, if we are dealing with a distribution of attributes,[23] it is always possible to calculate the values of

[19] Cambridge Philosophical Transactions, Vol. XX, 1904, pp. 36-65 and 113-141.

[20] Loc. cit.

[21] In statistical work the practice is followed of using the moments of the distribution for determining the parameters of the frequency curve. The i th moment μ_i of a frequency distribution about the arithmetic mean is by definition

$$\mu_i=\frac{\Sigma yx^i}{\Sigma y}$$

In calculating such moments it is necessary to consider the observations as grouped about the mid-point of the class interval and unless this interval is very small certain errors are introduced which can be partially eliminated by applying Sheppard's corrections as given by him in Biometrika, Vol. III, pages 308 seq. If Δx be taken as unity, we have

$$\mu_1=0 \qquad \beta_1^2=k=\frac{q-p}{\sqrt{pqn}}$$
$$\mu_2=pqn=\sigma^2$$
$$\mu_3=pqn(q-p) \qquad \beta_2=3+\frac{1-6pq}{pqn}$$
$$\mu_4=3(pqn)^2+pnq(1-6pq)$$

and if p is approximately equal to q and n is large we have $\sigma_k=\sqrt{\frac{6}{N}}$ and $\sigma_{\beta_2}=\sqrt{\frac{24}{N}}$.

[22] It is of interest to note that several investigators have derived this law independently. Thus H. Bateman derives this expression in an appendix to the article of Prof. Rutherford and H. Geiger previously referred to. This is, in a way, an illustration of the apparent need of a broader dissemination of information relating to the application of statistical methods of analysis to engineering and physical data. It is also of interest to note that this law has been used to advantage in the discussion of telephone trunking problems.

[23] If the classification is based upon the presence or absence of a single characteristic, this characteristic is often referred to as an attribute.

p, q and n from the moments of the distribution.[24] Even when p, q and n are known, the arithmetic involved in calculating the terms of the binomial is often prohibitive, and, therefore, it is necessary to obtain certain approximations corresponding to the three laws of error; that is, normal, second approximation, and the law of small numbers. Tables for the normal law and for the law of small numbers are readily available in many places, while those for the second approximation are given by Bowley.[25]

Even under conditions where the binomial expansion does not hold, Edgeworth has shown that it is possible to obtain the following general approximation:

$$y=\frac{1}{\sigma\sqrt{2\pi}}\left(\exp.-\frac{x^2}{2\sigma^2}\right)\left[1-\frac{k}{2}\left(\frac{x}{\sigma}-\frac{x^3}{3\sigma^3}\right)+\frac{k^2}{8}\left(-\frac{5}{3}+\frac{5x^2}{\sigma^2}-\frac{5x^4}{3\sigma^2}+\frac{x^6}{9\sigma^6}\right)+\left(\frac{\mu_4-3\mu^2}{8\sigma^4}\right)\left(1-\frac{2x^2}{\sigma^2}+\frac{x^4}{3\sigma^4}\right)\right]. \quad (7)$$

This holds providing the observations are influenced by a large number of causes, each of which varies according to some law of error but not necessarily to the normal law.

Gram-Charlier Series. Gram, according to Fisher,[26] was the first to show that the normal law is a special case of a more generalized system of skew frequency curves. He showed that the arbitrary frequency function $F(X)$ can be represented by a series of terms in which the normal law is the generating function $\phi\ (X)$. Thus

$$F(X)=c_0\phi(X)+c_1\phi'(X)+c_2\phi''(X)+\ \ldots \quad (8)$$

where c_0, c_1, c_2, etc., are constants which may be determined from the moments of the observed data. This series is similar to that already mentioned in the above equation (7) which Edgeworth has obtained in several different ways. This law is of interest from the viewpoint of either a physicist or an engineer in so far as it gives him a picture of the casual conditions consistent with an accepted theoretical curve. Thus, if either the causes of variation are within a certain degree not entirely independent, or the errors are not linearly aggregated, the observed frequency distributions may be expected to conform to an equation such as 8. This equation has been found to fit a much larger group of observed distributions than the normal law

[24] See footnote 26.

[25] See for example Pearson, K.—Tables for Biometricians and Statisticians—Cambridge University Press.

[26] Fisher, Arne—Theory of Probabilities—page 182.

and the publication of the necessary tables by Fisher [27] and Glover [28] makes the study of such a curve more feasible. The author finds, for example, that this series furnishes a much closer fit to the distribution of shots, Fig. 5, referred to above than any other that he has tried.

Theoretically we should be able to improve the approximation by taking a large number of terms of the series. Such a procedure, however, involves the use of moments higher than the first four, and the errors in these moments are so large as to make their use impractical.

In spite of the uncertainty attached to the interpretation of the physical significance of fitting any of these curves to data, one very practical observation has been made: that is, if an observed series of frequencies could not be fitted by a theoretical curve in any of the ways already mentioned, careful consideration of the possible reasons for the observed poor fit have in practically every instance suggested the cause or causes thereof. We shall refer to only one practical example.

The data have already been given above in Table II. It has been noted that in this instance the variations in the averages of groups of several thousand observations showed that the differences were significant. If the observed distributions had been normal, it would have been necessary to assume either that the methods of making the measurements were different for the different groups of observers, and for the different machines, or that the manufacturing methods were experiencing a trend. Although the observed frequency curves for the different groups were found to be smooth, the observed frequencies could not be readily fitted by any curve previously described. This naturally led to a search for the existence of any one of a number of causes affecting the observations which might produce such a divergence between theory and practice. One by one these causes were found and eliminated and as they were the degree of fit between the results of theory and practice increased. For example, it was found that some of the groups of observations were for transmitters assembled from only two or three lots of carbon. Transmitters assembled from one lot of carbon had a different average efficiency from those assembled from another lot. Naturally the

[27] Fisher, Arne—Loc. cit. As noted by Mr. Fisher, page 214, the values of $\phi(x)$ and its first 6 derivatives to 7 decimal places for values of x up to 4 and progressing by intervals of 0.01 were given by Jörgensen in his "Frekvensflader og Korrelation."

[28] Glover, J. W.—Tables of Compound Interest, Functions, etc.—1923 Edition published by George Wahr, Ann Arbor, Michigan.

resultant distribution was a compound of a few separate but similar distributions about different averages. When the distributions of the efficiencies of the different lots of carbon were determined separately they were found to be consistent with the second approximation.

Thus, although it may be impossible to conclude that the a priori assumptions underlying a given law of distribution are fulfilled because the observations are found to be consistent therewith, nevertheless, the fact that the observed and the theoretical distributions do not agree suggests the necessity of seeking for certain typical causes which may be expected to introduce such discrepancies. This point is of special importance in connection with the study of ways of sampling product in order to determine whether or not the manufacturing process is subject to trends. Thus, if a product is sampled at two periods, and the distributions of both groups of observations are found to be random about different averages, it is highly probable that the difference indicates a trend in the manufacturing methods, providing the difference between the averages is greater than 3 times the standard deviation of the average. When, however, the two distributions are found to be inconsistent with a random system of causes, it is quite probable that the condition of sampling has not been carefully controlled.

Hypergeometric Series. Pearson has shown several ways in which a frequency distribution may be represented by a hypergeometric series. Thus the chances of getting r, $r-1$, . . . 0 bad transmitters from a lot containing pn bad and qn good and where r instruments are drawn at a time may be represented by the terms of such a series. More important, however, is Pearson's solution [29] of what he calls the fundamental problem of statistics. He shows, following the line of reasoning similar to that originally suggested by Bayes, that if in a sample of $k_1=(m+n)$ trials, an event has been observed to occur m times and to fail n times, in a second group of k_2 trials the chances of the event occurring r times and failing s times are given by the successive terms of a hypergeometric series. We cannot consider here the questions underlying the justification of this method of solution, for, as is well-known, the application of Bayes' theorem is questioned by many statisticians. We can profit, however, by the broad experience of Prof. Pearson, for he has apparently accumulated an abundance of data which are consistent with the theory.

The answer to this problem s of special importance in connection with the inspection of product which in many instances runs into millions yearly. We must keep the cost of inspection at a minimum,

[29] Pearson, K.—Biometrika, October, 1920—pp. 1-16.

which means that the sample numbers must be small, and yet we see from the solution derived from Pearson the significance of the sizes of both the original and the second sample. Thus, he [30] shows that the standard deviation σ is given by the equation

$$\sigma^2 = k_2 pq\left(1 + \frac{k_2}{k_1}\right). \tag{9}$$

Multimodal Distributions. These occur frequently in engineering work and particularly in connection with the inspection of large quantities of apparatus. One such instance has already been referred to in the discussion of the data given in Table II, and another is illustrated by the data given in Fig. 1. Prof. Pearson [31] has developed a method for determining analytically whether or not the observed distribution is such as may be expected to have arisen from the combination of two normal components, the mean values of which are different. The method involves the solution of a ninth degree equation. As a result, the arithmetic work is in many cases prohibitive. This method cannot be applied to the data given in Fig. 1 primarily because the number of observations is not sufficiently great.

Pearson's Closed Type Curves.[32] One of the best known statistical methods for graduating data is that developed by Prof. Pearson. His system of closed type curves arises from the solution of the differential equation derived upon the assumption that the distribution is uni-modal and touches the axis when $y = 0$. In the hands of Pearson and his school great success has been attained in graduating data collected from widely different fields, although primarily from these of biology, psychology, and economics. The choice of curve to represent a given distribution rests primarily upon a consideration of a criterion involving two constants, $\beta_1 = \sqrt{k}$ and β_2, both of which have been defined previously in footnote 21.

In the early study of the distributions of efficiencies of product transmitters an attempt was made to apply this system of curves. For example, the Pearson types are indicated in Table II. In no instance, however, was it possible to obtain a very satisfactory fit between the observed and the theoretical distributions. Furthermore, the arithmetical work required to calculate a theoretical distribution in this way is excessive. We must also consider what physical significance can be attached to the different types of curves. The answer is not definite. Under certain conditions the generalized

[30] Pearson, K.—*Philosophical Magazine*—1907, pp. 365-378.

[31] Pearson, K.—*Philosophical Magazine*—Vol. 1, 1901, pp. 115-119.

[32] Elderton—Frequency Curves and Correlation.

equation of Pearson breaks down to the normal law and the second approximation. These, of course, can be explained as previously. The fundamental equation, however, serves to cover the condition where the causes are correlated. Thus, because of the lack of a clear conception of the physical significance of the observed variations in the type of curves indicated in Table II, it was not possible easily to set up experiments to find the causes of these variations. For this reason preference has been given to the use of frequency distributions derived upon a less empirical basis following the original lines laid down by La Place, Edgeworth, Kapteyn, and others previously referred to. Another very practical reason for choosing the latter type of curve is that it involves for the most part the use of only the first three moments of the distribution instead of the first four required for differentiating between the Pearson types. In those cases where the interest is less of physical interpretation than of graduating an observed set of data, preference may go to the more generalized system of Pearson.

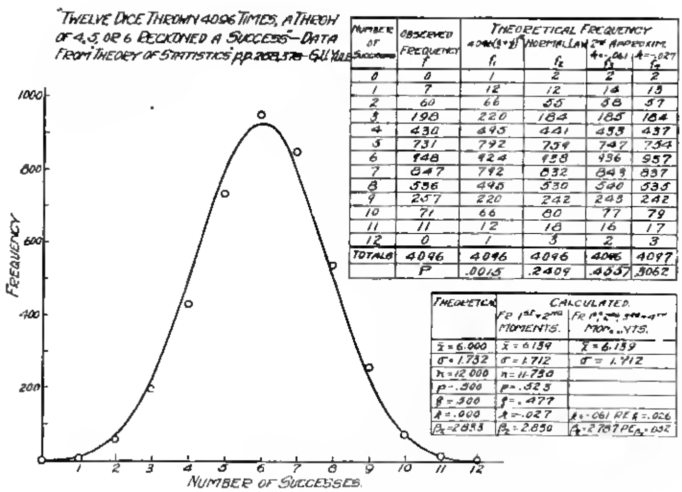
How Can We Choose the Best Theoretical Frequency Distribution?

We have already briefly reviewed some of the different methods for obtaining a theoretical frequency distribution from a consideration of the moments of the observed frequencies. We have seen in Table III that by using different methods we obtain different degrees of approximation to the hypothetically observed distribution which in this case corresponds to the terms of the binomial expansion $1000(.1+.9)^{100}$. Similarly from Fig. 5 it is seen that the Gram-Charlier series is a much closer approximation to the observed distribution than that derived upon the assumption of the normal law. In any given case we are naturally confronted with the question: What is the best theoretical distribution? We shall consider four methods for obtaining an answer.

The oldest, simplest, and in many instances the most practical, is that of comparing graphically or in tabular form the theoretical distribution with the one observed. This method is, however, inaccurate and qualitative. It does not furnish us with a quantitative method of measuring the closeness of fit between theory and practice, and in certain instances it is absolutely misleading. It is of interest to see how all of these things can be truly said of one and the same method. The first two characteristics, that is, oldest and simplest, are perhaps readily granted. It remains to be pointed out more

definitely wherein the method is sadly deficient as a quantitative measure, and therefore often misleading; whereas in certain instances it may be, nevertheless, the only practical method that can be used.

Graphical Method. The graphical method itself may be subdivided into two parts. Let us consider first the plot of the observed and theoretical frequencies. As an example of the unsatisfactory nature of this form of comparison, it is of interest to consider certain data



NUMBER OF SUCCESSES	OBSERVED FREQUENCY f	THEORETICAL FREQUENCY $4096(\frac{1}{2}+\frac{1}{2})^{12}$ f_1	NORMAL LAW f_2	2ND APPROXIM. $k=-.061$ f_3	$k=-.027$ f_4
0	0	1	2	2	2
1	7	12	12	14	15
2	60	66	55	58	57
3	198	220	184	185	184
4	430	495	441	453	437
5	731	792	754	747	754
6	948	924	958	936	957
7	847	792	832	849	837
8	536	495	530	540	535
9	257	220	242	245	242
10	71	66	80	77	79
11	11	12	18	16	17
12	0	1	3	2	3
TOTALS	4096	4096	4096	4096	4097
	P	.0015	.2409	.4557	.5062

THEORETICAL	CALCULATED FR. 1ST & 2ND MOMENTS.	FR. 1ST, 2ND, 3RD & 4TH MOMENTS.
$\bar{x}=6.000$	$\bar{x}=6.139$	$\bar{x}=6.139$
$\sigma=1.732$	$\sigma=1.712$	$\sigma=1.712$
$n=12.000$	$n=11.750$	
$p=.500$	$p=.523$	
$q=.500$	$q=.477$	
$k=.000$	$k=-.027$	$k=-.061$ P.E.$_k=.026$
$\beta_2=2.833$	$\beta_2=2.850$	$\beta_2=2.787$ P.E.$_{\beta_2}=.032$

Fig. 8

given by Yule[33] in which 12 dice are thrown 4,096 times, a throw of 4, 5, or 6 points being reckoned a success. If the dice are symmetrical $p=q=\frac{1}{2}$ and the theoretical distribution if given by 4,096 $(\frac{1}{2}+\frac{1}{2})^{12}$, the terms of which as given by Yule are presented in the third column of Fig. 8. It is suggested that the reader, before going further, consider the graphical and tabular representation of these data. The smooth curve is the theoretical distribution $4{,}096(\frac{1}{2}+\frac{1}{2})^{12}$. It has been the author's experience to find that in practically every instance in which this curve has been shown to an individual for the first time that the impression is that which Yule evidently desires to produce by the illustration: that is there is a very good fit between theory and practice. This distribution is, however, not symmetrical: it is skew. The dice used in this experiment were not symmetrical: that is, $p\neq q$. How do we know that these statements are true?

Let us consider the normal and second approximation as given

[33] Yule—"Introduction to the Theory of Statistics."

in the fourth, fifth, and sixth columns.[34] Obviously the degree of fit is closest for the second approximation, although that between the normal distribution and the observed frequencies is closer than that between the terms of the binomial expansion and the observed frequencies. To be sure, the normal law is only an approximation to the point binomial when $p=q$ and $n=\infty$. The normal distribution, however, is calculated about the observed average 6.139, instead of about the theoretical average 6. If the dice are non-symmetrical, the average will not be 6, and, therefore, the center of the distribution will be shifted after the fashion observed. The improvement in fit corresponding to the normal distribution is therefore primarily attributable to that introduced by shifting the center of the distribution indicating that $p \neq q$. However, if $p \neq q$, the second approximation should improve the fit and for either value of k this is found to be the case. Thus even though we cannot measure quantitatively the improvement of fit, the qualitative evidence presented in this figure is sufficient to warrant the conclusion that the dice were non-symmetrical, and therefore, that the smooth curve is an unsatisfactory graduation of the data. In fact, by using a quantitative method for measuring the goodness of fit to be discussed in a succeeding paragraph, it follows that only 15 times out of 10,000 can we expect a divergence from theory as large or larger than that exhibited by the frequencies corresponding to the point binomial.

We have also previously called attention to the fact that in Fig. 7 the eye does not serve to differentiate satisfactorily between the distribution calculated upon the assumption of the normal law and that given by the binominal expansion when the conditions underlying the normal law are far from being satisfied.

Regardless of these criticisms, such graphical methods cannot be entirely dispensed with. Thus the graphical representation of the data given in Fig. 1 shows very clearly that the distribution is probably bimodal, although with no more observations than are available it is practically impossible to show that this is true in any other way.

Instead of plotting the frequency y of occurrence of a variable of magnitude x as ordinate, and x as abscissa, the practice is often followed of plotting as ordinate the percentage of the total number N of observations having magnitudes of x or less.[35]

Any curve $\phi(y, x)=0$ may be replaced by a straight line.[36] In

[34] Two values of k were calculated as indicated in the lower right hand corner of the figure.

[35] Heindlhoffer, K. and Sjövall, H.—Endurance Test Data and their Interpretation—Advance paper presented at the Meeting of the American Society of Mechanical engineers, Montreal, Canada, May 28 to 31, 1923.

[36] Runge, C.—*Graphical Methods,* p. 53.

this way we can transform the integral curve into a straight line by choosing an x-scale proportional to the integral from 0 to x of the probability curve.[37] When plotted in this way, a normal distribution appears as a straight line on such paper. At first it may appear very simple to determine whether or not the data conform to a straight line, but in practice this is not always so easy. Thus, we have seen that the distribution of shots presented in Fig. 5 is not normal, but

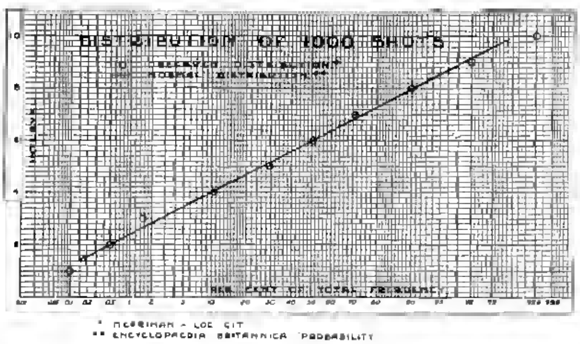


Fig. 9

when these results are plotted on probability paper we have the curve given in Fig. 9. The reader should be cautioned that in such a case there is a temptation to consider that the observed points are approximately well fitted by the straight line, although this is not the case.

Probability paper could be ruled for different theoretical distributions, but in its present form it serves only to determine whether or not the distribution is approximately normal. Its use leaves much to be desired in the way of a quantitative measure of the degree of fit between the theoretical and observed distributions.

Calculation of σ, $\beta_1 = \sqrt{k}$, *and* β_2. Let us consider what information can be obtained as to the best theoretical distribution from only a consideration of the first four moments of the observed frequencies. Let us consider the values of k and β_2 presented in Table V. These have been calculated for the point binomial $(p+q)^n$ where p, q and n have been given different values. For the normal law corresponding to $p=q$ and $n=\infty$, we have $k=0$ and $\beta_2=3$. Thus, if in a practical

[37] Whipple, G. C.—The Elements of Chance in Sanitation—*Franklin Institute Journal*, Vol. 182, July, December, 1916—pp. 37-59 and 205-227.

TABLE V

p	$n=4$		$n=9$		$n=16$		$n=25$		$n=100$		$n=10{,}000$	
	k	β_2	k	β_2	k	β_2	k	β_2	k	β_2	k	β_2
.5	0	2.50	0	2.78	0	2.87	0	2.92	0	2.98	0	3.00
.6	−.20	2.54	−.14	2.80	−.10	2.89	−.08	2.93	−.04	2.98	−.004	3.00
.7	−.44	2.69	−.29	2.86	−.22	2.92	−.17	2.95	−.09	2.99	−.009	3.00
.8	−.75	3.06	−.50	3.03	−.38	3.02	−.30	3.01	−.15	3.00	−.015	3.00
.9	−1.33	4.28	−.69	3.57	−.67	3.32	−.53	3.20	−.27	3.05	−.027	3.00
.99	−4.92	26.75	−3.28	13.56	−2.46	8.94	−1.97	6.80	−.98	3.95	−.098	3.01
.999	−15.79	251.75	−10.52	113.45	−7.90	65.19	−6.31	42.76	−3.16	12.95	−.316	3.10
.9999	−50.00	2501.75	−33.33	1113.44	−25.00	627.69	−20.00	402.76	−10.00	102.94	−1.000	4.00

case we find an observed distribution for which $k=0$ and $\beta_2=3$, it is highly probable that the distribution is approximately normal. It is true, however, that in sampling from a universe in which $p=q$ and $n=\infty$, the observed values of k and β_2 will seldom be exactly equal to 0 and 3 respectively. Then we must ask what range of values may be expected in these two factors for distributions which are practically normal. For such cases the variations in k and β_2 are practically normal [38] and have standard deviations $\sigma_k=\sqrt{\frac{6}{N}}$ and $\sigma_{\beta_2}=\sqrt{\frac{24}{N}}$ where N is the number of observations. Thus, theoretically any series of observations for which the calculated values of k and β_2 fall within the ranges $0\pm 3\sigma_k$ and $3\pm 3\sigma_{\beta_2}$ may have arisen from a normal universe. Since, however, the errors σ_k and σ_{β_2} of sampling are so large, this method does not furnish a very practical test for distribution consisting of only a few observations. This is particularly true since, even for very skew distributions, the values of k and β_2 do not differ much from 0 and 3 respectively (see Table V). If, however, the number of observations is large, the values of k and β_2 in themselves often indicate very definitely that the observed frequencies are not consistent with the normal law. For example the calculated values of k and β_2 given for the inspection data in Table II show conclusively that in practically every instance the observed data could not have arisen from a normal universe. So long as we do not use Pearson's system of curves, all that these two factors indicate is that the observed data do or do not conform to the normal law and in this respect their use is limited as is that of the probability paper mentioned above.

In order to show that the factor β_2 is not in itself a very sensitive measure of the variability from the normal law, I have considered the following special case. Let us assume that the observed distributions can be grouped into two parts depending upon whether or not the observations cluster about the average $\overline{X}_1$ or $\overline{X}_2$ measured from a point which is the arithmetic mean of the entire distribution taken about a common origin. This corresponds to the practical case such as that indicated by Fig. 1 which as already pointed out often occurs in practice.

[38] For a critical study of the conditions under which the probable errors of these constants have a real significance, reference should be made to a discussion of this problem by Isserlis in the Proceedings of the Royal Society, series A, Vol. 92, pp. 23 seq.—1915. Obviously even for the normal distribution all of the moments will be skew. This follows from a consideration of equation 4.

The value of β_2 for the entire distribution is then given by the following expression:

$$\beta_2 = \frac{(\overline{X}_1{}^4 \sum y_1 + \overline{X}_2{}^4 \sum y_2) + 6(\overline{X}_1{}^2\sigma_1{}^2 \sum y_1 + \overline{X}_2{}^2\sigma_2{}^2 \sum y_2)}{(\sum y_1 + \sum y_2)\mu_2{}^2}$$

$$+\frac{4(\overline{X}_1\ {}_1\mu_3 \sum y_1 + \overline{X}_2\ {}_2\mu_3 \sum y_2) + ({}_1\mu_4 \sum y_1\ {}_2\mu_4 \sum y_2)}{(\sum y_1 + \sum y_2)\mu_2{}^2},$$

where ${}_1\mu_i$ and ${}_2\mu_i$ refer to the adjusted ith moments of the observations about their respective mean values. Let us assume that $\overline{X} = \overline{X}_1 = \overline{X}_2$; $k_1 = k_2 = 0$; ${}_1\beta_2 = {}_2\beta_2 = 3$; ${}_1\mu_i = {}_2\mu_i$; $\sum y_1 = \sum y_2$; and $\sigma_1 = \sigma_2$ where $\sum y_1$ and $\sum y_2$ represent the total numbers of observations in the first and second groups respectively. It may be shown by substitution in this equation that, if $|\overline{X}| = |\sigma_1|$, $\beta_2 = 2.5$, whereas, if $|\overline{X}| = |10\sigma_1|$, $\beta_2 = 1$, approximately. Thus, if the numbers of observations in each of the two sub-groups are the same and the component curves are normal, the value of β_2 for the entire distribution about the mean of the two will, in general, decrease as $|\overline{X}|$ becomes large in comparison with $|\sigma_1|$. Differences in β_2 of this magnitude are difficult to establish. Furthermore the skewness is zero, and therefore does not indicate the bi-modal character of the distribution.

Let us consider the case where $|a\,\overline{X}_1| = |\overline{X}_2|$; $k_1 = k_2 = 0$; ${}_1\beta_2 = {}_2\beta_2 = 3$; $\sum y_1 = a \sum y_2$; ${}_1\mu_i = {}_2\mu_i$. If, $a = 10$ and $|\overline{X}_1| = |\sigma_1|$ then $\beta_2 = 8+$ whereas if $|\overline{X}_1| = |10\sigma_1|$, then $\beta_2 = 100$, approximately.[39] Thus, for comparatively wide differences in the averages, it requires a large number of observations in order to increase the precision of β_2 to such an extent as to prove the significance of deviations in this factor of the magnitudes noted above.

The skewness in this case is not zero and its significance could be established with a comparatively small number of measurements. In any of the above cases a carefully constructed plot would serve to indicate the bimodal characteristic of the curve better than the study of the factor β_2.

Pearson's Criterion of Goodness of Fit. A much more powerful

[39] Here again it should be noted that the values of β_2 are independent of the actual frequencies of each of the two groups and depend only upon the ratio of these frequencies and upon the ratio of $|\overline{X}_1|$ to σ_1.

criterion has been developed by Prof. Pearson [40] in a series of articles in the *Philosophical Magazine.* It is true that this test for goodness of fit cannot be used indiscriminately. In fact the application of this criterion is subject to numerous limitations clearly set forth in the original papers by Pearson and in more recent articles on the mathematics of statistics. In the use of the method it is necessary that these be kept in mind by the individual making the original analysis of the data. Irrespective of these facts, however, the method itself is one of the most useful tools available for measuring in a quantitative way the "goodness of fit" between two distributions. The significance of the values of P given in Figs. 5, 6, and 8 now become evident.

Engineering Judgment. The fourth very practical and one of the most useful methods of comparing the theoretical with the observed distribution is that of applying common sense or engineering judgment. To quote from a recent article of Prof. Wilson [41] we have: "And as the use of the statistical method spreads we must and shall appreciate the fact that it, like other methods, is not a substitute for, but a humble aid to the formation of a scientific judgment." Even with the use of all the statistical methods known to the art, it remains impossible to determine the true nature of the complex of causes which control a set of observations. We can present plausible explanations, but we can never be sure that they are right. Sometimes we can present two plausible explanations and then we must fall back on engineering judgment or common sense to decide between them. A striking illustration of this fact is presented in the following paragraph.

Prof. Pearson [42] has recently presented measurements of the cephalic index of a certain group of skulls. The object of the investigation was to determine if variation had gone on to such an extent as to indicate the survival of the fitter inside a homogeneous population, or the survival of two races both of which were in existence many ages in the past. Pearson shows that, by a solution of a nonic equation,

[40] If we divide the entire range of variation into s equal intervals for which the observed frequencies are $f_1, f_2, \ldots f_s$ and the corresponding theoretical frequencies are $f_1', f_2', \ldots\ldots f_s'$, Pearson calculates the function

$$\chi^2 = \sum \frac{(f'-f)^2}{f'}$$

from which he is able to determine the probability that a series of deviations as large as, or larger than, that found to exist could have arisen as a result of random sampling. Tables have been prepared which give the probability of fit in terms of the number of intervals into which the entire range has been divided and of the value of χ.

[41] Wilson, E. B.—The Statistical Significance of Experimental Data—science—New Series, Vol. 58, 1493, October 10, 1923, pp. 93-100.

[42] *Philosophical Magazine,* Vol. 1, 1901—pp. 110-124.

he is able to find two component distributions which when added together approximate very closely to the observed frequencies. The observed data are given in the second column of Table VI and the frequencies of Prof Pearson's compound curve are given in the third column of the table. The probability of fit between these two distributions is seen to be approximately .96, which is indeed very

TABLE VI

Rowgrave Skulls *

Cephalic Index	Observed Distribution f	Compound Distribution f_1	2nd Approximation f_2	$\frac{(f_1-f)^2}{f_1}$	$\frac{(f_2-f)^2}{f_2}$
67	1	1	1	0	0
68	1	2	2	.50	.50
69	3	4	4	.25	.25
70	8	7	8	.14	0
71	13	11	14	.36	.07
72	13	18	22	1.39	3.68
73	33	28	30	.89	.30
74	36	39	39	.23	.23
75	49	50	48	.02	.02
76	59	59	55	0	.29
77	69	65	59	.25	1.69
78	70	66	60	.24	1.67
79	54	60	58	.60	.28
80	58	52	53	.69	.47
81	40	43	46	.21	.78
82	31	35	39	.46	1.64
83	25	28	32	.32	1.53
84	28	23	26	1.09	.15
85	21	20	21	.05	0
86	20	17	16	.53	1.00
87	9	14	13	1.79	1.23
88	10	11	10	.09	0
89	6	8	7	.50	.14
90	10	6	5	2.67	5.00
91	2	4	3	1.00	.33
92	3	2	2	.50	.50
93	2	1	1	1.00	1.00
94	1	1	1	0	0
95	0	0	1		1.00
Σ	675	675	676	15.77	23.75
Probability of fit P				.957	.694

Ave. $=\bar{x}=78.846$	Ave. $=\beta_2=3.181$	Ave. $=\sigma_k=$ .0943
$\sigma=$ 4.612	$\sigma_{\bar{x}}=$ .178	$\sigma_{\beta_2}=$ .189
$k=$ .521	$\sigma_{\sigma}=$ .126	

* Phil. Mag., Vol. I, 1901, pp. 115–119.

high, meaning, of course, that 96 times out of 100 we may expect to find a system of deviations as large or larger than that actually found. The author finds, however, that the theoretical distribution

(column 4) based upon the assumption of the second approximation is also a very close fit to the observed frequencies, the probability of fit being in this case .69. As a result of these calculations shall we conclude that the distribution is composed of two normal components as indicated in Fig. 10, or shall we conclude that the distribution is homogeneous? In other words, do the skulls belong to two or to only

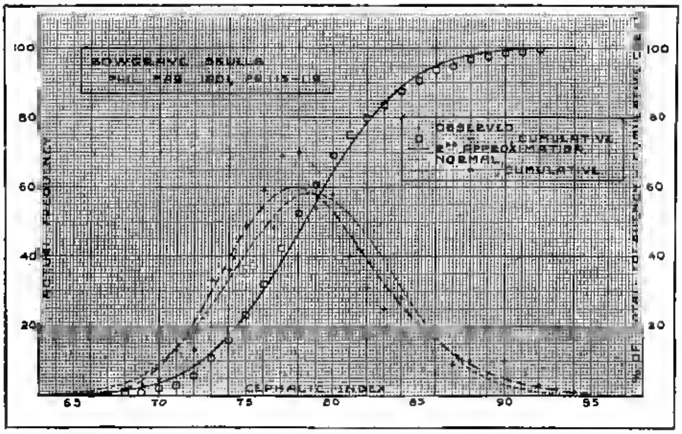


Fig. 10

one race? The measure given by the probability of fit is, of course, in favor of the first alternative. It is highly probable, however, that if we had been given the observed distribution without any discussion of what it meant we would have decided that it probably was consistent with the assumption of the random system of causes such as might underlie the second approximation.

In other words, if we had been given merely the above set of skull measurements, it is reasonable to suppose we might have concluded that the distribution was homogeneous. However, when our judgment is colored by the facts which cannot be presented in the array of observed frequencies we must conclude that it is highly probable that the observed data have arisen from a non-homogeneous population.

Statistical methods alone do not answer all of the questions that are raised in this problem nor do they answer them in many others. There is almost always room for judgment to enter.

Thus, analyzing a group of measurements of some characteristic of a large number of transmitters, it often becomes necessary to determine whether or not they can be subdivided into normal com-

ponents as in the above problem. In our case the subgroups correspond to different kinds of carbon. Here, as in the data given by Pearson, it often has been found necessary to base our final conclusion partly upon facts not revealed by the data themselves.

The integral curves corresponding to the normal and observed distributions are given in Fig. 10 in order to show that they do not

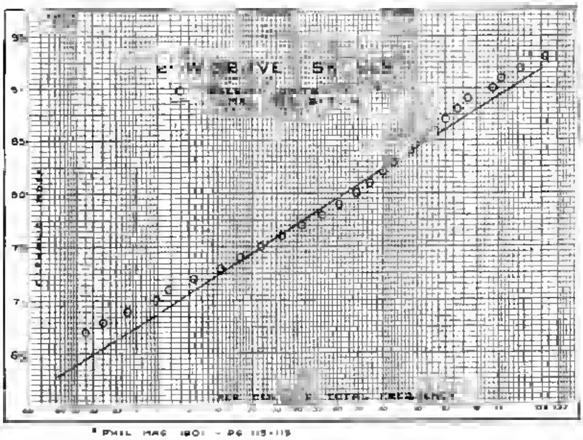


Fig. 11

serve to indicate the difference between the observed and theoretical distributions nearly as well as the actual frequency curves also given in this figure. Fig. 11 presents the result on probability paper. In this case the probability curves are as good as the frequency curves for showing the divergence between theory and observation. It will be recalled that this is not true for the similar curves given in Fig. 9.

Summary Statement of Suggested Method to be Followed in the Analysis of Engineering and Physical Data

We have briefly reviewed the different methods for determining the best theoretical distribution to represent observed data. The following four steps indicate the ordinary procedure:

1. Obtain the first four corrected moments.
2. Calculate the average, standard deviation, k and β_2, and their standard deviations.
3. Calculate the theoretical distribution of distributions warranted by the circumstances.

4. Apply one or more of the four methods of comparing the theoretical and observed distributions to determine which one is theoretically the best.[43]

An illustration of the method of applying this form of analysis to inspection data on transmitters is indicated in the schematic chart Fig. 12. The object of the inspection of apparatus in the process

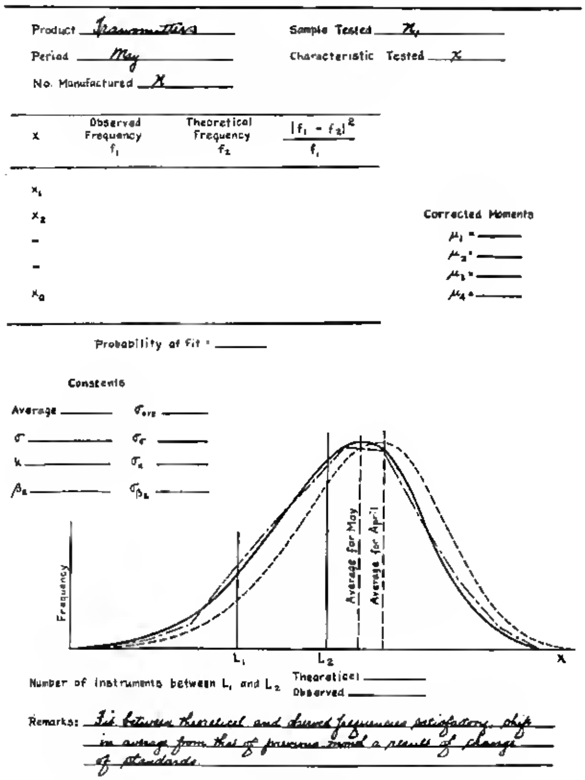


Fig. 12

of manufacture is obviously to determine the most probable law of distribution, and from this to determine whether or not there is any indication of a trend in the quality of the product. In the light of what has been said, it is obvious that a complete report of this character should contain the items called for in Fig. 12. The corrected

[43] If the observed distribution could not have arisen from a random system of causes, it may be advisable to attempt to transform it into an approximately random one, such as was done in connection with the data in Fig. 6.

moments and the factors, such as the average, standard deviation, k and β_2 should be given. These factors provide us with measures of the lack of symmetry, and can be used as pointed out in the previous sections of this paper. Recording this amount of data makes it possible for anyone interested, either to check the calculations of the theoretical frequencies and the conclusions derived therefrom, or to calculate a different theoretical distribution based upon fundamentally different hypotheses in a way such as has been illustrated already in the discussion of the distribution of measurements of the cephalic index, as given in Fig. 11.

In most instances, however, it is highly probable that the man who originally prepares the chart is charged with the responsibility of choosing the best distribution, and, therefore, the chief interest of those reading the report is centered upon the conclusions indicated therein. The graphical representation of the observed distribution by means of the histogram is hopeful. The comparison of this with the theoretical curve represented by a solid line shows qualitatively whether or not the product is changing. The probability of fit gives a quantitative measure of the degree of fit. The set of curves given in Fig. 12 is drawn to illustrate a condition which may sometimes happen when, for example, the standards used in the machines have been changed. This is only typical of the results which may be expected. Obviously, the form of such reports designed to meet specific conditions will vary. That presented above is only typical of one which has been found to be of value in presenting the analysis of the results of inspection of certain types of apparatus.

Some Advantages Derived from a Comparatively Complete Statistical Analysis

It has been pointed out that the value of either a physical or an engineering interpretation of data depends upon the success attained in deriving the best theoretical distribution. This is the equation which fits the observed points best, and which, if possible, can be interpreted physically. The previous discussion indicates the way in which different causal relationships tend to produce typical frequency distributions, and also the way in which statistical methods may be used in finding a theoretical distribution which yields a physical interpretation.

This point has been illustrated by several examples. It has been shown that by a proper choice of theoretical curve a very close approximation to an observed distribution can be obtained. This

TABLE VII—FREQUENCY DISTRIBUTION OF σ PARTICLES*

Number of Particles	Observed Frequency f	†Normal Law f_1	2nd Approximation f_2	‡Law of Small Numbers f_3	§Poisson Charlier f_4	$(p+q)^n$ f_5	$\frac{(f-f_1)^2}{f_1}$	$\frac{(f-f_2)^2}{f_2}$	$\frac{(f-f_3)^2}{f_3}$	$\frac{(f-f_4)^2}{f_4}$	$\frac{(f-f_5)^2}{f_6}$
−3	0	1	0	0	0		1.000				
−2	0	5	0	0	0		5.000				
−1	0	23	8	0	0		23.000	8.000			
0	57	74	63	54	47	50	3.905	.571	.167	2.128	.180
1	203	180	197	210	196	202	2.939	.183	.233	.250	.005
2	383	337	389	407	400	405	6.279	.093	1.415	.123	1.195
3	525	485	532	525	533	533	3.299	.092	0	.120	.120
4	532	535	526	508	525	521	.017	.068	1.134	.093	.232
5	408	452	398	394	407	402	4.283	.251	.497	.002	.090
6	273	294	249	254	258	255	1.500	2.313	1.421	.872	1.271
7	139	146	139	140	138	137	.336	0	.007	.007	.027
8	45	57	72	68	64	64	2.526	10.125	7.779	5.641	5.641
9	27	15	28	29	25	26	9.600	.036	.140	.160	.038
10	10	4	10	11	9	9	9.000	0	.091	.111	.111
11	4	1	3	4	3	3	9.000	.333	0	.333	.333
12	**0	0	1	1	1	1	∞	1.000	1.000	1.000	1.000
13	**1	0	0	0	0	0					
14	**1	0	0	0	0	0					
Σ	2608	2609	2615	2605	2606	2608	81.684	23.065	13.884	10.615	10.245
Probability of Fit							.0000	.0410	.3086	.5624	.5946

$p=.046$

$q=.954$

$\bar{x}=pn=3.87$

$\sigma=\sqrt{pqn}=1.92$

$K=\frac{q-p}{\sqrt{pqn}}=.48$

$n=84.174$

$\beta_2=3+\frac{1-6\,pq}{pqn}=3.51$

$\sigma_k=\sqrt{\frac{6}{2608}}=.0480$

$\sigma_{\beta_2}=\sqrt{\frac{24}{2608}}=.0959$

($N=64$ used in computing $(p+q)^n$)

* Prof. Rutherford and H. Geiger—Phil. Mag., Oct., 1910.

† Normal calculated on basis of probability integral.

‡ Bateman, H., in appendix to original article.

§ Fisher, A.—loc. cit., p. 273,

** These observed frequencies are grouped together in computing P.

has already been indicated in Table III. To emphasize this point, however, let us consider once more the distribution of alpha particles given in Table I. These data together with various theoretical [44] distributions are given in Table VII.

Let us consider the data given in Table I by following the procedure of analysis outlined in the previous section. The factors k and β_2 when compared with their errors should indicate whether or not the distribution is normal. As shown in Table VII, k and β_2 differ from 0 and 3 respectively, by more than 3 times their respective standard deviations. As has already been pointed out, this is sufficient evidence to indicate that the distribution is not normal. In order to show, however, that if we follow the next step and calculate theoretical distributions based upon the assumption of the different laws; that is, in this case, normal, second approximation, and the law of small numbers, we are naturally led to the choice of the best distribution. This choice is materially influenced by the measure of the probability of fit as recorded in the table. The law of small numbers is obviously a very close approximation to the observed frequencies.

One of the obvious things to do in this problem, but one that has not been done previously, is to calculate the values of p, q and n, and from them the terms of the binomial expansion $2608(p+q)^n$. The probability of fit between the terms of this expansion and the observed frequencies is the highest given in the table. This increases the evidence that the distribution is random. It also does nore. It serves to establish the facts that the probability p that an alpha particle will strike the screen is .046, and that the maximum number of alpha particles which may ever be expected to strike the screen is of the order of magnitude of 84. Granted then that we can always find the most probable theoretical frequency distribution, let us consider next the influence that the result may have in our determination of the most probable value, the number of observations between any two limits and the casual relationships governing the distribution.

Let us consider first the dependence of the most probable value upon the type of distribution. In our present work in the study of carbon the resultant distributions have been in most instances either random or such that through a proper transformation they could be reduced to such. For any distribution consistent with the second approxima-

[44] The source of all distributions previously calculated are indicated. The Poisson-Charlier series is similar to the Gram-Charlier series, except that the law of small numbers is the generating function. It serves as an admirable method of graduating certain classes of skew distribution as illustrated by this example and by that given in Table III.

tion the most probable value is at a distance $-\frac{k\sigma}{2}$ from the arithmetic mean. Many distributions have been found for which k lies between .5 and unity, and, therefore, this difference is from ¼ to ½ of the standard deviation. Thus, the efficiencies of certain standard types of transmitters are found to conform to such a law, and the difference between the modal and average values is of the order of magnitude of 0.4 mile.

Obviously the geometric mean of the sound intensities (Fig. 6) and not their arithmetic mean is the most probable. The difference between the two is quite large. The difference between the arithmetic mean and the modal value for groups of data such as given in Fig. 1, Tables II and VI are quite large. To use again the illustration ot the alpha particles the observed most probable number is 4; whereas, the observed average[45] is 3.87. Judging from the best theoretical distribution the most probable number of alpha particles is 3. Choosing the number 3 it is seen that either of the other two numbers differ from this by approximately ½ the standard deviation. Such results are, however, not confined to the work of the present investigation nor to the examples previously cited as is evidenced by the data given in the last column of Table VIII.

TABLE VIII

N=Number of Observations	Source of Data	Percentage Within $\overline{X}\pm\sigma$	Percentage Within $\overline{X}\pm 2\sigma$	Percentage Within $\overline{X}\pm 3\sigma$	$\frac{\text{Average} - \text{Modal}}{\sigma}$
1000	E *54	66.6	97.2	99.6	.803
251	E 66	78.1	94.8	97.6	1.042
9154	E 10	67.7	95.5	99.6	.031
2162	E 79	70.1	95.1	99.3	−.311
368	E 84	73.4	94.6	97.0	.422
675 Table VI.........		68.7	94.1	99.6	.247
Normal Law......		64.26	95.44	99.73	0

* Elderton "Frequency Curves and Correlation," published by C. & E. Layton, London, 1906.

We should not leave this phase of the discussion, however, without pointing out that in a large number of purely physical experiments a sufficient number of observations has not been taken to make it possible to choose the best theoretical distribution. In general more than

[45] Of course, such an average has no significance, except for a continuous distribution.

100 observations are required. Thus, in Prof. Millikan's [46] determination of the electron charge e only 58 observations were made. The values of σ, k, and β_2 for this distribution are .128 units, $-.196$ and 2.358. Even though the observed distribution is consistent with a normal system of causes, values of k and β_2 may be expected to occur which differ from 0 and 3 respectively, as much as these observed values do. In this case even if k is real and not a result of random sampling, the correction to be added to the average in order to obtain the most probable value is insignificantly small.

Next let us consider the problem of determining the number of observations between any two limits. The physicist is ordinarily concerned with the probable error: that is, the error such that ½ of the observations lie within the range $\overline{X}\pm$ probable error. Its magnitude for the normal distribution is $.6745\sigma$, and the errors are distributed symmetrically on either side of the average. It is interesting to note that the magnitude of the probable error is also $.6745\sigma$ for the second approximation, but that the errors are not distributed symmetrically on either side of the average.

Another important pair of limits is that including the majority of the observations. For the normal law 99.73% of the observations are included within the range $\overline{X}\pm 3\sigma$ which, therefore, is often called the range. Not a single example has been found, however, of a distribution for which the observed number of observations within this range is less than 95% even though the distribution is decidedly skew. In fact it is seldom less than 98%. If, however, we have a case such as that represented in Table II where groups of observations have been taken in what is technically known as different universes, and then averaged together, the average result is not the most probable, and the standard deviation of the average is not inversely proportional to the square root of the number of observations. Since this point is of considerable importance, it is perhaps well to state it in a slightly different way. Thus, let us assume that we have a thousand samples of granular carbon which possess inherent microphonic efficiencies differing by comparatively large magnitude. Transmitters assembled from any one of the groups of carbon cover a range of efficiencies. If we choose a sample of 10,000 instruments, 5,000 from each of two lots of carbon which do not possess the same inherent efficiency, we cannot expect, for reasons already pointed out, that the observed distribution will be normal. The average of these observations will not in general be the most probable value, and the standard deviation of the average will not be equal to the

[46] Millikan, R. A.—The Electron—*University of Chicago Press.*

observed standard deviation divided by the square root of the number of observations, in this case 10,000.

We have already seen, however, that it is possible to detect such errors of sampling, since in general the distribution cannot be fitted by the second approximation or Gram-Charlier series. If the theoretical distribution is either normal, second approximation, or the law of small numbers, the number of observations to be expected between any two limits can be readily determined from the tables. Experience has shown that in every instance where it has been possible to represent the observed distribution in any of these three ways, the data obtained in future samplings have always been consistent with the results to be expected from the theory underlying these three laws. It will be of interest to note the data given in columns 3, 4, and 5 of Table VIII and to compare the theoretical percentages (last row) for the different limits with those observed.

In closing it is of interest to point out further the significance of some of the results discussed in this paper in connection with the inspection of equipment. Here we must decide upon a magnitude of the sample to be measured in order to determine the true percentage of defective instruments in the product. If p is the percentage defective, and q that not defective, then the standard deviation about the average number found in a sample of n chosen from N instruments is

$$\sigma^2 = pqn\left(1-\frac{n}{N}\right).$$

In practice, however, we never know the true value of p unless we measure all of the apparatus, and this is impractical. In our calculations we must therefore use some corrected value. We find, though, that the average value of p is in most instances the one that must be used. Assuming that we choose a value of p, the distribution of defectives in N' samples of n in number will be represented by the distribution of $N'(p+q)^n$. If one of the samples is found to contain a percentage of defectives, which is inconsistent, that is, which is highly improbable as determined from the distribution of $N'(p+q)^n$, it indicates that the product is changing.

If, however, we take into account the effect of the size of the first sample in respect to the second as indicated by Pearson,[47] we see that the distribution of N' samples may be different from that given by the binomial expansion. In accordance with this theory, if in a first sample of 100, 10% of the sample is found to possess a given attribute,

[47] Pearson, K. Loc. cit. Foot note 30.

the distribution of the percentages to be expected in 1,000 such samples is indicated by the last column of frequencies in Table III. In order to show graphically how this distribution differs from that

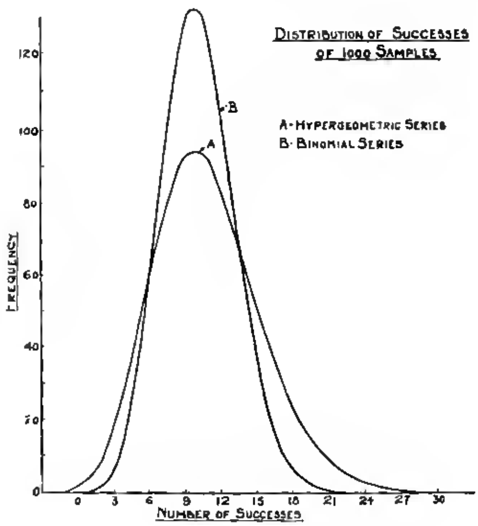


Fig. 13

corresponding to the binomial expansion these two sets of frequencies are reproduced in Fig. 13. The difference between them is a striking illustration of the significance of the size of the samples used in connection with the inspection of equipment, providing we accept Pearson's results.